CDF note 10133

Search for a low mass Standard Model Higgs boson in the di-tau decay channel with 2.3 fb^{-1} of CDF data.

The CDF Collaboration
URL http://www-cdf.fnal.gov

(Dated: July 14, 2010)

We present the result of a search for the Standard Model (SM) Higgs boson in the $\tau\tau$ decay mode, using approximately 2.3 fb^{-1} of CDF Run II data, collected at the Tevatron $p\bar{p}$ collider at $\sqrt{\mathrm{s}} = 1.96$ TeV. The search is performed considering the following signal processes: WH($\rightarrow \tau\tau$), ZH($\rightarrow \tau\tau$), qHq'$\rightarrow$q$\tau\tau$q'and gg$\rightarrow$H$\rightarrow \tau\tau$.

Events are selected by requiring an hadronic tau and one isolated electron or muon, coming from the leptonic decay of one of the two taus. In addition, at least one calorimeter jet must be present in the final state.

We expect 921.8±48.9 background events in the 1 jet channel and 159.4±11.6 in the $\geq$ 2 jets channel, while in data we observe 965 and 166 events, respectively.

In order to improve the search sensitivity we employ a multivariate technique, based on a set of Boosted Decision Trees trained to get the best separation between signal and the dominant sources of background. We observe no evidence for a Higgs boson signal and therefore we set a 95% confidence level (C.L.) upper limit on the cross section relative to the SM predictions ($\sigma/\sigma_{\mathrm{SM}}$). Results are presented for the Higgs boson mass varying from $\mathrm{M_H} = 100$ GeV/c^2 to $\mathrm{M_H} = 150$ GeV/c^2. For the mass hypothesis of 120 GeV/c^2 the observed limit is 27.2, while the corresponding expected value is $23.4^{+9.8}_{-6.4}$.

I. INTRODUCTION

In the Standard Model the mechanism which provides masses to fundamental particles via electroweak symmetry breaking predicts the existence of a neutral scalar particle, the Higgs boson. The mass of the Higgs boson is a free parameter in the SM; however, direct searches at LEP experiments have set a lower limit of 114.4 GeV/c^2 at 95% C.L., while precision electroweak measurements constraint its value to be less than 157 GeV/c^2 [1].

The dominant production mode at the Tevatron collider is gluon-gluon fusion which proceeds via a virtual top quark loop, while the main decay process, for masses below 135 GeV/c^2, is into a $b\bar{b}$ quark pair. Because of the overwhelming QCD background in the gg→H→ $b\bar{b}$ channel, the Tevatron experiments have focused their searches on the Higgs produced in association with a Z or a W boson, where the Higgs subsequently decays into b quarks.

The analysis presented in this note explores the complementary decay mode into a pair of tau leptons. Although the branching ratio of this process is smaller by one order of magnitude, a lot of sensitivity is recovered by including four different production channels:

- WH($\rightarrow \tau\tau$)
- ZH($\rightarrow \tau\tau$)
- qHq'→q$\tau\tau$q'
- gg→H→ $\tau\tau$

The analysis is performed using 2.3 fb^{-1} of data collected by the CDF detector. We select the events by requiring that one of the two taus decays leptonically, while the other one is reconstructed by its hadronic decay products.

A multivariate technique, based on a set of the Boosted Decision Trees (BDTs) [2], has been adopted to provide a more performing tau identification (ID) algorithm, with respect to the CDF standard cut-based selection.

The request of at least one additional calorimeter jet in the final state increases the sensitivity for vector boson fusion events and for the Higgs associated production with a W or Z boson decaying hadronically. In this analysis we require exactly two identified leptons; thus leptonic decays of W and Z are admitted only when the additional leptons coming from them are lost or fail the identification requirements. Acceptance for Higgs coming from gluon fusion is small, but signal yield is comparable to the other processes because of the much higher production cross section.

The sensitivity of the search is further enhanced by employing another set of BDTs, trained to discriminate the Higgs signal from the principal sources of background in events containing one or more than two calorimeter jets in the final state. Since no evidence of a Higgs signal is found, a bayesian method is used to calculate the 95% C.L. upper limit on the production cross section, based on the combination of the BDT templates defined for the two signal channels.

II. DETECTOR DESCRIPTION

The CDF II detector is a cylindrically simmetric spectrometer surrounded by calorimeters and muon detectors. The geometry is described using the azimuthal angle φ and the pseudorapidity $\eta = -\ln\tan(\vartheta/2)$, where ϑ is the polar angle with respect to the proton beam axis. A complete description of the detector can be found elsewhere [3], here we briefly present only the subsystems relevant for this analysis.

The charged particle tracking system consists of a set of silicon micro-strip detectors [4] [5], covering the region $|\eta|$ <2, surrounded by a 96-layer drift chamber ($|\eta|$ <1) [6] immersed in a 1.4 T solenoidal magnetic field, parallel to the incoming particle beams. Outside the solenoid, sampling calorimeters segmented with projective tower geometry are used to reconstruct electromagnetic showers and hadronic jets. The calorimeters are divided into a central ($|\eta|$ <1.1) and a forward region (1.1< $|\eta|$ <3.6). A set of strip and wire chambers, embedded in the calorimeter at the approximate electromagnetic shower maximum (SMX), determines the shower shape and the centroid positions and helps in discriminating photons from electrons.

A system of drift chambers and scintillation counters surround the calorimeters and are devoted to the detection of muons. The integrated luminosity of $p\bar{p}$ collisions is measured by Cherenkov luminosity counters.

III. BOOSTED DECISION TREE: KEY TOOL OF THE ANALYSIS

The fundamental tool implemented in this search is represented by the Boosted Decision Tree multivariate analysis method.

A *decision tree* is an event classifier based on a sequence of rooted binary splits, performed using a set of discriminating variables. Given a training sample made of known signal and background events, repeated decisions are performed: at each *node* the variable and the split value which give the best separation are selected and two classes (or child nodes) are created. When a predefined criterium is met, the splitting sequence stops and the terminal nodes, called *leaves*, are tagged as signal (S) or background (B) according to the purity of the content; a score of +1 or -1 is then assigned to each event depending on the leaf type.

The *boosting* consists in the creation of a *forest* of trees: training events which were misclassified in one tree, have their weight increased in the subsequent one. For each event the final score is finally given by the average of the different tree outputs. The boosting procedure stabilizes the response of the decision trees and makes the algorithm more robust and less sensible to fluctuations in the samples.

Once a trained BDT is applied to a sample with an unknown composition of signal and background, two different approaches may be adopted to benefit from the discriminating power of the algorithm:

- by simply placing a lower cut on the classifier output response, a fraction of the events is retained, with a higher efficiency for the signal component with respect to the background;
- by keeping the whole output distribution, all the information coming from the shapes can be exploited to discriminate signal against background.

The first approach is employed in the definition of the new tau ID algorithm, while the second one is adopted for the final sensitivity optimization of the analysis. A more detailed description will be provided in section IV and VIII, respectively.

IV. LEPTON IDENTIFICATION

Tau leptons are very short lived particles, which can be detected by CDF only through their decay products. Hadronic decays, denoted as τ_{had} have a branching ratio of about 65% and their signature is characterized by very narrow jets in the calorimeter, matching reconstructed tracks (mainly charged pions) and π^0's. The remaining channels are represented by the leptonic decays into muons (τ_μ) or electrons (τ_e). Depending on the combinations of these modes, several possible final states are defined for the H$\rightarrow \tau\tau$ search, with different branching ratios and background contributions. In this analysis we consider only the $\tau_h\tau_e$ and $\tau_h\tau_\mu$channels, corresponding to the 46% of all possible combinations.

Events are collected with a dedicated trigger which requires one electron or muon candidate, with transverse momentum $\text{p}_\text{T} > 8$ GeV/c and one additional isolated track with $\text{p}_\text{T} > 5$ GeV/c, which is used as a starting point for the hadronic tau reconstruction.

Electrons and muons are fully reconstructed off-line and must satisfy several quality requirements. In particular, electron candidates need to be associated to a well-measured track and must have an energy cluster consistent with originating from an electromagnetic shower. Muons are identified by a charged track releasing a small amount of energy in the calorimeter and matched to a reconstructed track segment in the muon chambers. In addition, both the electron and the muon have to be isolated such that the sum of the transverse energies for the calorimeter elements in a cone of $\Delta\text{R} = \sqrt{\Delta\varphi^2 + \Delta\eta^2}$ <0.4 around the lepton is smaller than 10% of the electron transverse energy E_T or muon p_T.

The hadronic tau system is identified by building a signal cone ($\theta < \theta_{sig}$) around a good quality seed track having $\text{p}_\text{T} > 6$ GeV/c and satisfying the trigger level isolation requirements . Because of the Lorentz boost, the collimation of the tau decay products increases with the original energy of the tau; as a consequence, θ_{sig} is defined as a function of the energy of the calorimeter cluster of the candidate (E_{clu}), i.e. $\theta_{sig} = \min(0.17, 5.0\ \text{GeV}/\text{E}_{\text{clu}})$ rad.

Since the neutrino is not detected, only a partial reconstruction of the tau momentum is possible: charged and neutral pions are used to build the so called "visible" four-momentum, defined as $p^{\text{vis}} = \Sigma p^{\text{track}} + \Sigma p^{\pi^0}$. On the other hand, by demanding no additional charged tracks or π^0s in the isolation region $\theta_{sig} < \theta < \theta_{\text{iso}}$ outside the signal cone (where $\theta_{\text{iso}} = 0.52$ rad) it is possible to suppress the background represented by jets originating from guarks or gluons.

This search is based on the requirement of only one (1-prong) or three (3-prong) tracks with $\text{p}_\text{T} > 1$ GeV/c in the signal cone, such that the resulting charge is ± 1. Instead of the common CDF cut-based selection, we apply a different identification strategy: candidates are selected by means of a set of Boosted Decision Trees, which are trained to discriminate real hadronic tau decays against QCD jets, by exploiting all the information available in the signal and in the isolation regions.

Dedicated BDTs are defined for several tau subcategories, accordingly to the visible energy and the track multiplicity of the candidates; a lower cut is set for each BDT output distribution, in order to keep the identification efficiency at the same level of standard CDF selection and to maximize the jet fake rejection.

Finally, also electrons and muons, which could mimic the tau signature by depositing large amounts of energy in the electromagnetic calorimeter are removed, by applying a cut on $\left[\frac{E_{\rm tot}}{\rm P} \times \left(1 - \frac{E_{\rm em}}{E_{\rm tot}}\right)\right] > 0.2$, where P, $\rm E_{tot}$ and $\rm E_{em}$ are the scalar sum of the momenta of all tracks in the signal cone, the total and the electromagnetic cluster energies, respectively.

V. EVENT SELECTION

The general strategy of this analysis is to apply very simple and minimal requirements to keep Higgs acceptance as high as possible, and then rely on a multivariate technique to exploit all the kinematical and topological information to discriminate signal against the principal sources of background.

The final selection consists of exactely one central isolated electron(muon) with offline $\rm E_T(p_T)>$ 10 GeV(/c) and one central hadronic tau with visible $\rm p_T >$ 15 GeV/c. The two reconstructed leptons must have opposite charges, need to be spatially well separated (ΔR >0.4) and come from the same interaction point. In addition, the good quality reconstructed primary vertex is required to be in the region consistent with the beam-beam interaction.

In order to further clean up the event candidate selection and increase the signal purity, we then apply the following requirements:

- Electrons which are likely originating from photon conversions are removed;
- Events which come from cosmic rays interactions in the detector are rejected;
- Events with tau candidates which are likely muons with a large deposite of energy in the calorimeter due to brehmstrahlung, are rejected;
- Events in the $\tau_e\tau_{had}$ case, where the hadronic tau is a 1-prong with $E_{\rm had}/P < 0.4$, and the invariant mass of the dilepton system lies between 80 and 110 GeV/c^2, are vetoed to suppress Z→ee contribution.

Calorimeter jets are clustered with a fixed cone size of ΔR < 0.4 and are required to be well separated from the identified leptons, have an electromagnetic fraction smaller than 0.9, an $\rm E_T$, corrected for instrumental effects [16], greater than 20 GeV and a $|\eta|$ smaller than 2.5.

The missing transverse energy $\not\!\!E_T$, defined as the vector sum of all the calorimeter tower energy depositions projected on the transverse plane, is used to identify the presence of neutrinos in the final state. The $\not\!\!E_T$ is corrected for the presence of muons, when their calorimeter deposit is different from the track-derived momenta, and for the corrections applied to the energy of all jets with raw $\rm E_T >$ 10 GeV and $|\eta|$ <2.5.

VI. BACKGROUND AND SIGNAL ESTIMATE

The processes which contribute to the expected event yield of our analysis selection are modeled by Monte Carlo (MC) simulations and data-driven techniques.

We determine the geometrical and kinematical acceptances of the signal and of the main sources of physics irreducible backgrounds, including $Z/\gamma^* \to \tau\tau$, $Z/\gamma^* \to \mu\mu$/ee, $t\bar{t}$, WW, WZ and ZZ, by using samples of simulated collision events: Higgs, top-antitop and diboson processes are generated with PYTHIA MC [8], while multiparton final states for Drell-Yan plus jets events are estimated using ALPGEN [9], matched with PYTHIA for the hadronization and parton showering. The initial state momenta of partons are modeled by using CTEQ5L parton distribution functions (PDF's [10]). A GEANT-based run dependent simulation [7] models the interaction of all final particles in the detector subsystems.

The expected number of events for each MC-derived processes is obtained by considering the corresponding theoretical cross sections and by scaling to the data sample luminosity covered by this search; correction factors are properly applied on an event by event basis, for the effects due to the trigger requirements, the z-vertex position and for the lepton reconstruction efficiencies, as summarized in the following formula:

$$\mathrm{N}^i = \sigma^i \times \mathrm{A}^i \times \varepsilon_{\rm trig} \times \varepsilon_{\rm ID} \times \varepsilon_{\rm vtx} \times \int \mathrm{Ldt} \quad (1)$$

where i refers to the considered background process, σ is the corresponding cross section, $\varepsilon_{\rm trig}$ is the trigger efficiency given by the product of the tau and the other lepton (electron or muon) trigger efficiency, $\varepsilon_{\rm ID} = \varepsilon_{\rm IDtau} \times \varepsilon_{\rm IDlep}$ is the resulting ID scale factor for the two identified leptons, $\varepsilon_{\rm vtx}$ is the run dependent efficiency of the z vertex position requirement, $\int$ Ldt is the luminosity used in the analysis, A^i is the acceptance for the process under investigation.

The tau BDT-based identification efficiency MC correction factor has been evaluated by comparing data to the background expectations in the $Z/\gamma* \rightarrow \tau\tau$ control region. A data-based procedure is applied for the estimation of the following background processes, coming from misidentified leptons:

- γ + jet, where the photon undergoes conversion and fakes an electron, and the jet is misidentified as a hadronic tau;
- QCD multijet, where one jet fakes a tau and another one fakes an electron or a muon;
- W$\rightarrow$ lν+jets, where one of the jets in the final state fakes the hadronic tau.

In the first two cases no correlation is expected between the charges of the two reconstructed lepton candidates; as a consequence, the number of opposite sign (OS) events ($Q_{\rm lep} \times Q_{\tau}$= -1) is approximately equal to the number of same sign (SS) events ($Q_{\rm lep} \times Q_{\tau}$= 1). The latter can therefore be used to provide an estimate of γ+jet and QCD multijet fake background.

Also W+jets events give a contribution to the SS data samples; however, in this case a clear correlation appears between the charge of the lepton coming from the W boson and the charge of the outgoing quark, which generates the jet faking the hadronic tau: this results in an excess of OS events. This W+jets extra contribution, which needs to be added to the total background prediction, is estimated by comparing data to ALPGEN MC expectation in a W+jets enriched control region, obtained by loosening the tau ID requirements and by applying a cut on $\not{E}_T > 25$, 30, 35 GeV (for the 0 jet, 1 jet and $\geq$ 2 jets cases respectively) and on the transverse mass of the lepton and the $\not{E}_T$, M_T(lep,$\not{E}_T$)> 40 GeV/c^2.

VII. CONTROL SAMPLES AND SIGNAL REGIONS

We classify the selected events by the number of identified jets in the final state: the "1-jet" and "$\geq$ 2 jets" channels are studied separately and used for the Higgs signal search, while events with no jets are used to build several control samples, useful to test the background normalization and modeling. These control samples are orthogonal one from each other and are defined by the following requirements:

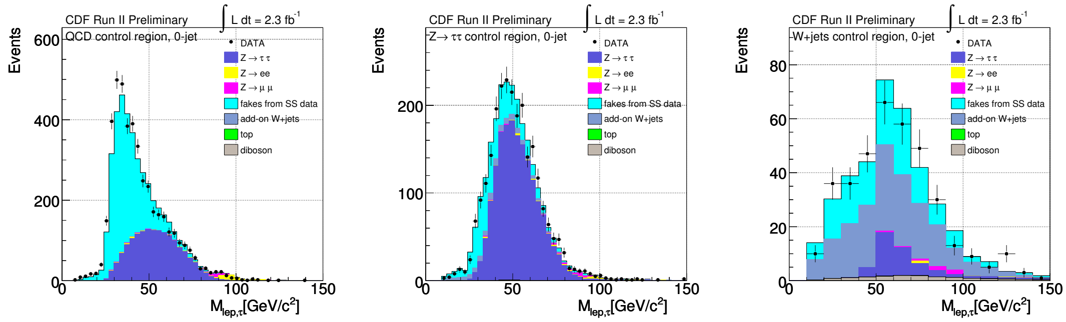
- QCD region: $\not{E}_T \leq 10$ GeV
- $Z/\gamma^* \rightarrow \tau\tau$ region: $\not{E}_T \geq 10$ GeV & $M_T(\rm lep, \not{E}_T) \leq 60$ GeV/c^2
- W+jets region: $\not{E}_T \geq 10$ GeV & $M_T(\rm lep, \not{E}_T) \geq 60$ GeV/c^2

In each control region we find a good agreement between the background estimations and the observed number of events, as summarized in table I. In figure 1 and 2 the invariant mass of the lepton (electron or muon) and hadronic tau system and the visible mass of the hadronic tau, for the three control regions, are reported.

The expected background, the signal yield and the observed data for the 1 jet and $\geq$ 2 jets signal samples are summarized in table II, while figures 3 and 5 show some significant kinematics and tau identification distributions.

Background source	QCD C.R.			$Z/\gamma^* \rightarrow \tau\tau$ C.R.			W+jets C.R.		
$Z/\gamma^* \rightarrow \tau\tau$	1610.4	±	72.2	1735.5	±	73.0	42.0	±	1.9
$Z/\gamma^* \rightarrow ee/\mu\mu$	120.6	±	5.4	44.0	±	2.0	6.9	±	0.3
WW/WZ/ZZ	1.2	±	0.1	9.7	±	0.8	14.9	±	1.2
$t\bar{t}$	0.011	±	0.002	0.10	±	0.02	0.24	±	0.05
fakes from SS data	2542.0	±	254.2	594.8	±	59.5	137	±	13.7
add-on W+jets	13.3	±	0.7	112.2	±	5.6	180.1	±	9.0
Total Background	4287.8	±	203.9	2496.3	±	54.9	381.4	±	14.9
Data		4433			2501			386	
Total Signal (M_H = 120 GeV/c^2)	0.63	±	0.11	1.253	±	0.203	0.126	±	0.016

TABLE I: Event yield in the 0-jet control regions, with 2.3 fb^{-1} of CDF data.



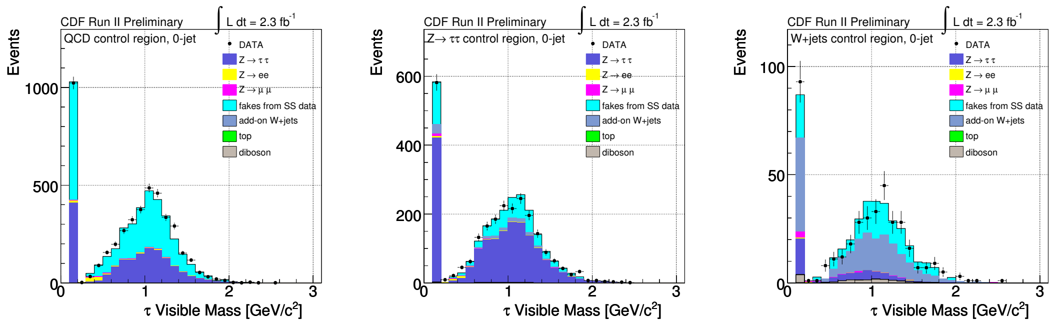
(a) QCD control region: invariant mass of lepton-tau system

(b) Z/$\gamma*\rightarrow\tau\tau$ control region: invariant mass of lepton-tau system

(c) W+jets control region: invariant mass of lepton-tau system

FIG. 1: data-MC comparison in the 0-jet control regions.

]h!]



(a) QCD 0jet control region: visible mass of the reconstructed hadronic tau

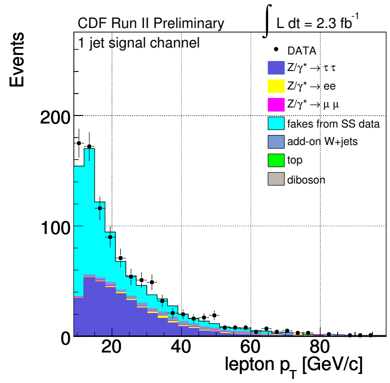
(b) Z/$\gamma*\rightarrow\tau\tau$ control region: visible mass of the reconstructed hadronic tau

(c) W+jets control region: visible mass of the reconstructed hadronic tau

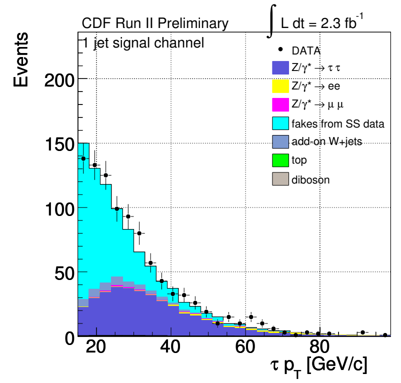
FIG. 2: data-MC comparison in the 0-jet control regions.

Signal channels M_H = 120 GeV/c^2						
Background source	1 JET			$\geq$ 2 JETS		
Z/$\gamma^*\rightarrow\tau\tau$	357.9	±	33.1	59.3	±	8.8
Z/$\gamma^*\rightarrow$ ee/$\mu\mu$	26.4	±	2.0	4.8	±	0.7
WW/WZ/ZZ	3.9	±	0.4	0.9	±	0.1
$t\bar{t}$	4.6	±	0.6	16.3	±	1.9
fakes from SS data	483.0	±	48.3	64.0	±	6.4
add-on W+jets	45.8	±	8.2	14.1	±	4.2
Total Background	921.7	±	48.9	159.4	±	11.6
Data		965			166	
ggH	0.535	±	0.154	0.129	±	0.092
WH	0.091	±	0.010	0.150	±	0.014
ZH	0.050	±	0.005	0.099	±	0.009
VBF	0.070	±	0.009	0.099	±	0.013
Total Signal	0.746	±	0.163	0.477	±	0.121

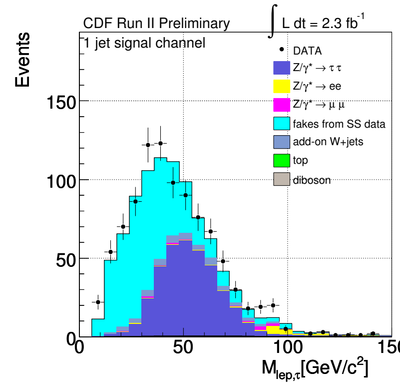
TABLE II: Event yield in the 1 jet and $\geq$ 2 jets channels, with 2.3 fb^{-1} of CDF data.



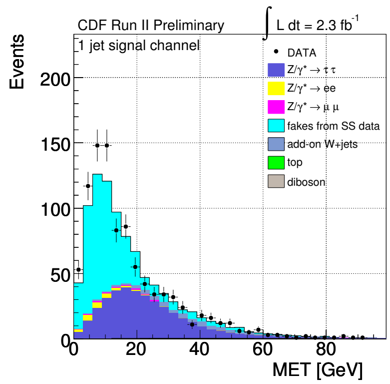
(a) Lepton(e/μ) p_T



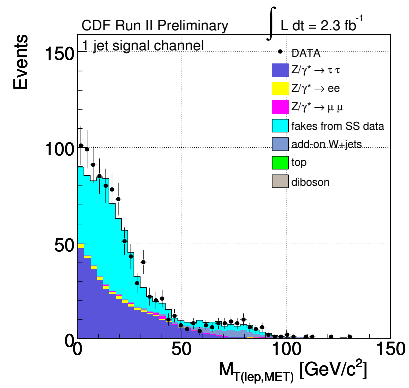
(b) Tau p_T



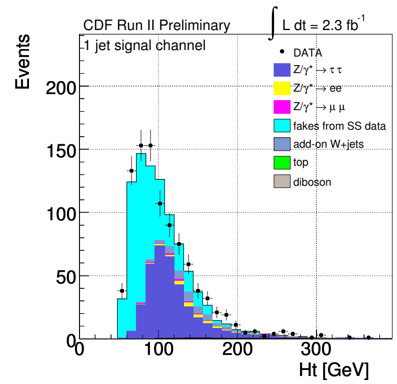
(c) Invariant mass of the lepton-tau system



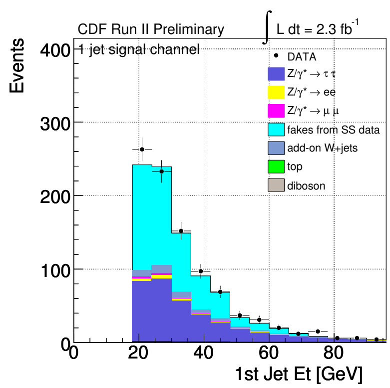
(d) Missing transverse energy



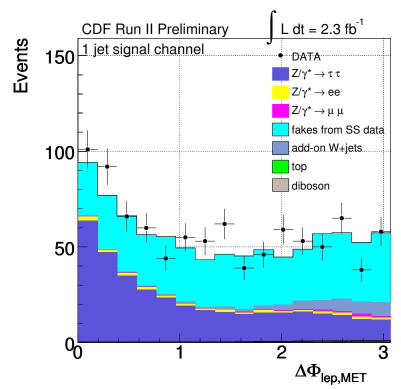
(e) Transverse mass of the lepton-$\not{E}_T$ system



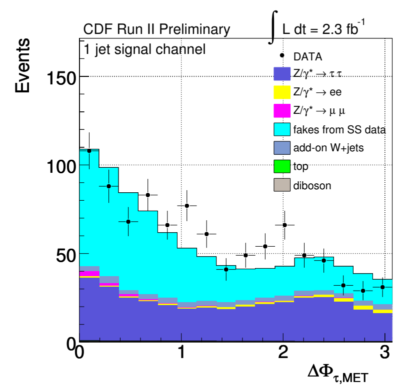
(f) ΣE_T of the reconstructed objects in the event



(g) Leading Jet E_T



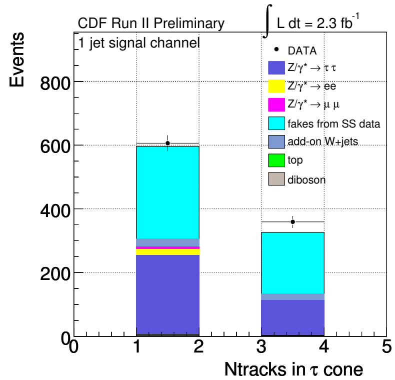
(h) $\Delta\varphi$ between lepton and $\not{E}_T$



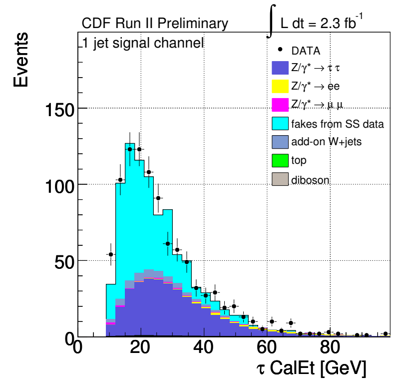
(i) $\Delta\varphi$ between tau and $\not{E}_T$

FIG. 3: data-MC comparison in the 1-jet signal channel: kinematic variables

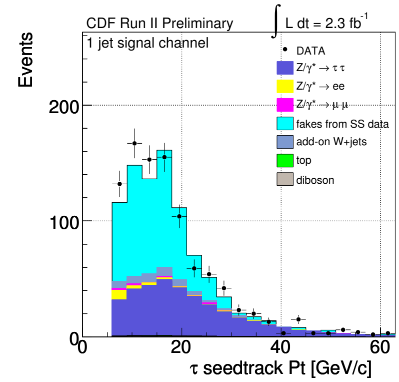
.



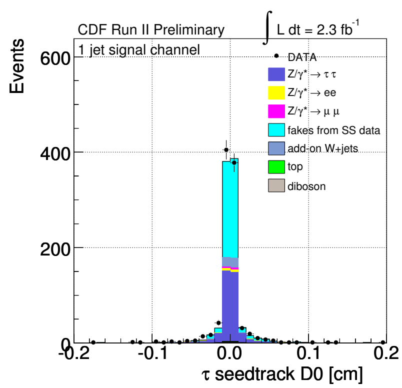
(a) Number of tracks(p_T >1 GeV/c) in tau signal cone



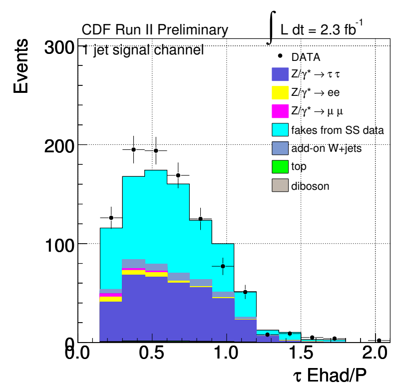
(b) Tau calorimeter E_T



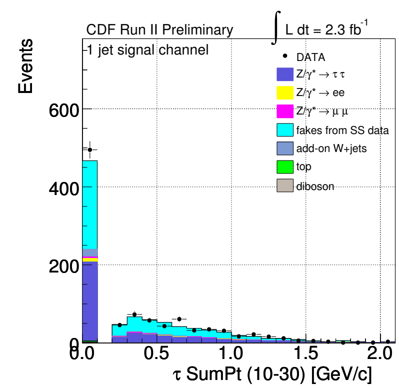
(c) Tau seed track p_T



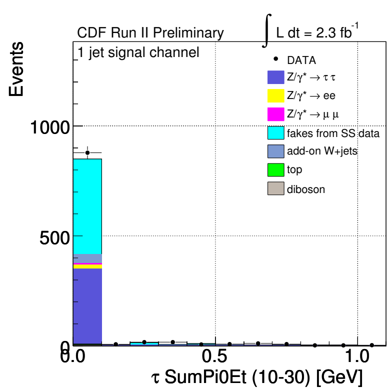
(d) Tau seed track impact parameter



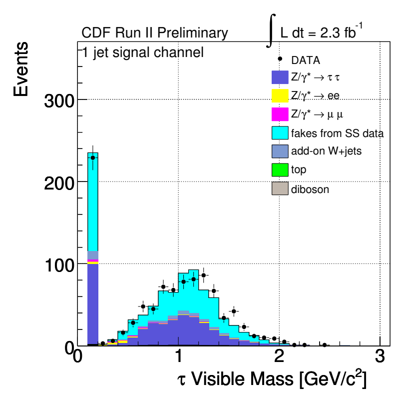
(e) Tau $E_{had/P}$



(f) Sum of p_T of all tracks in the tau isolation cone



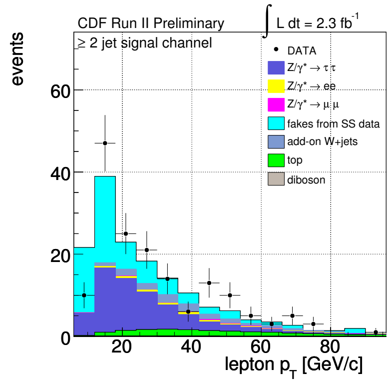
(g) Sum of E_T of all π^0's in the tau isolation cone



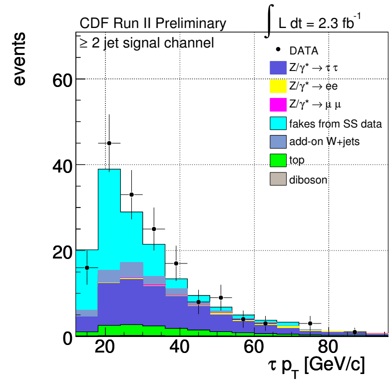
(h) Tau visible mass

FIG. 4: data-MC comparison in the 1-jet signal channel: hadronic tau variables

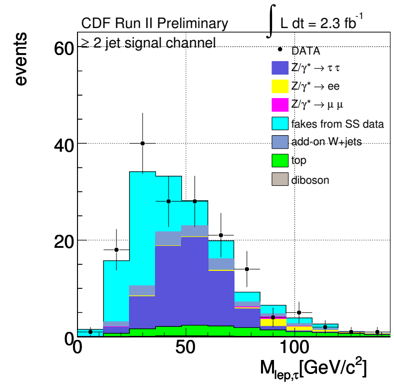
.



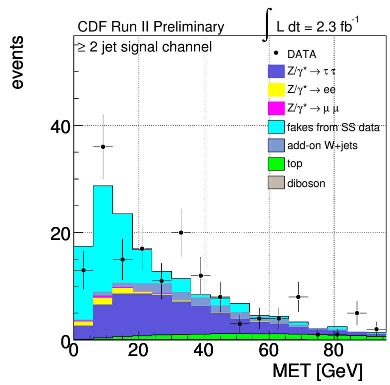
(a) Lepton(e/μ) $\mathrm{p_T}$



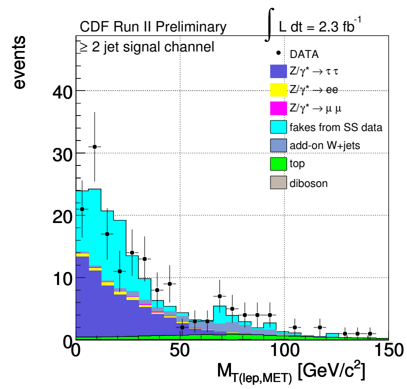
(b) Tau $\mathrm{p_T}$



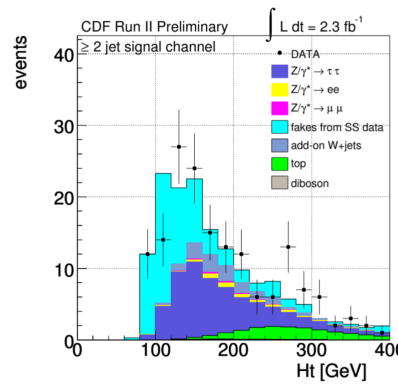
(c) Invariant mass of the lepton-tau system



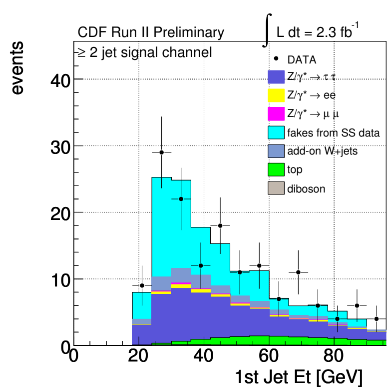
(d) Missing transverse energy



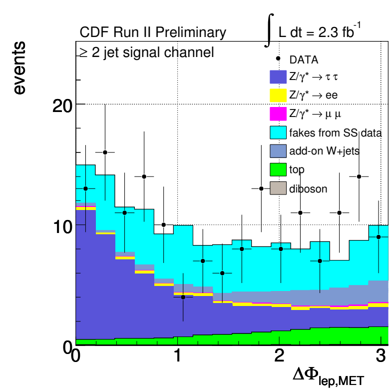
(e) Transverse mass of the lepton-$\not{E}_T$ system



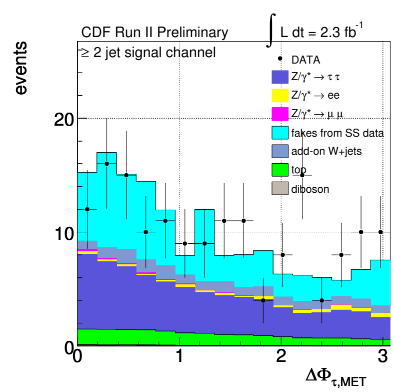
(f) ΣE_{T} of the reconstructed objects in the event



(g) Leading Jet $\mathrm{E_T}$

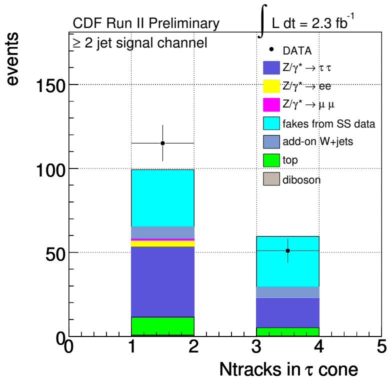


(h) $\Delta\varphi$ between lepton and $\not{E}_T$

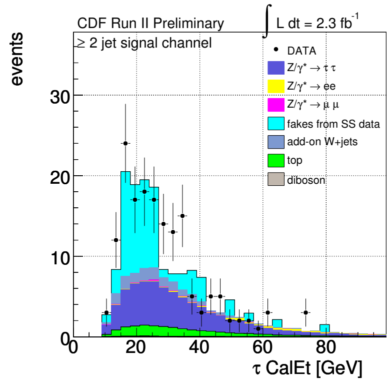


(i) $\Delta\varphi$ between tau and $\not{E}_T$

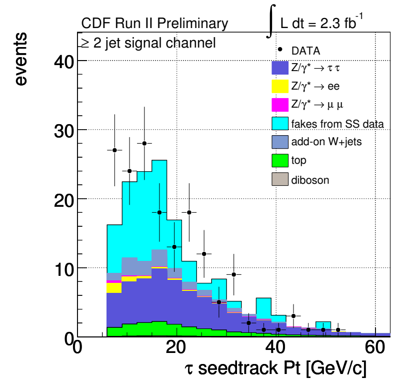
FIG. 5: data-MC comparison in the $\geq$ 2 jets signal channel: kinematic variables.



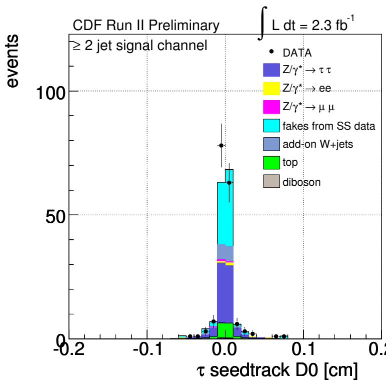
(a) Number of Tratks(p_T >1 GeV/c) in tau signal cone



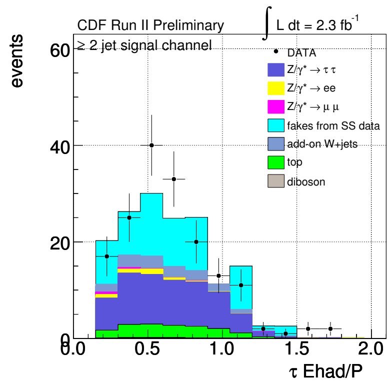
(b) Tau calorimeter E_T



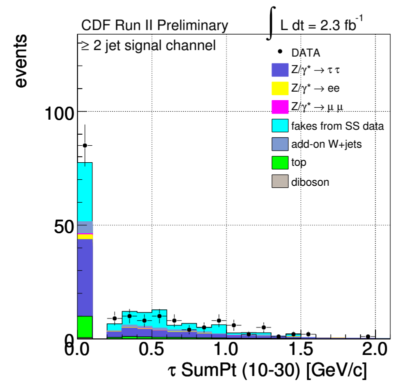
(c) Tau seed track p_T



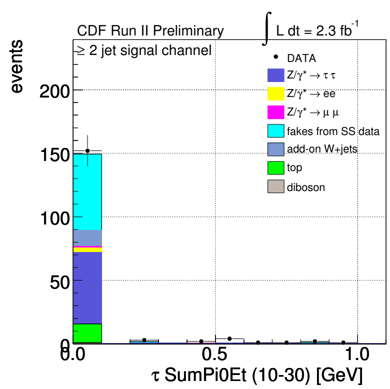
(d) Tau seed track impact parameter



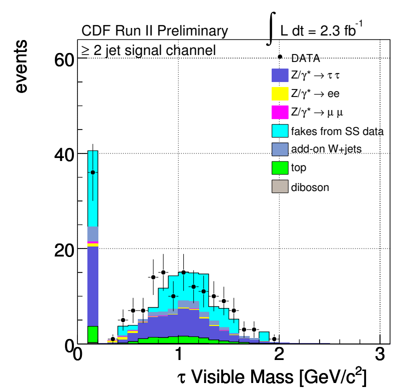
(e) Tau $E_{had/P}$



(f) Sum of p_T of all tracks in the tau isolation cone



(g) Sum of E_T of all π^0's in the tau isolation cone



(h) Tau visible mass

FIG. 6: data-MC comparison in the ≥ 2 jets signal channel: hadronic tau variables.

VIII. SENSITIVITY OPTIMIZATION

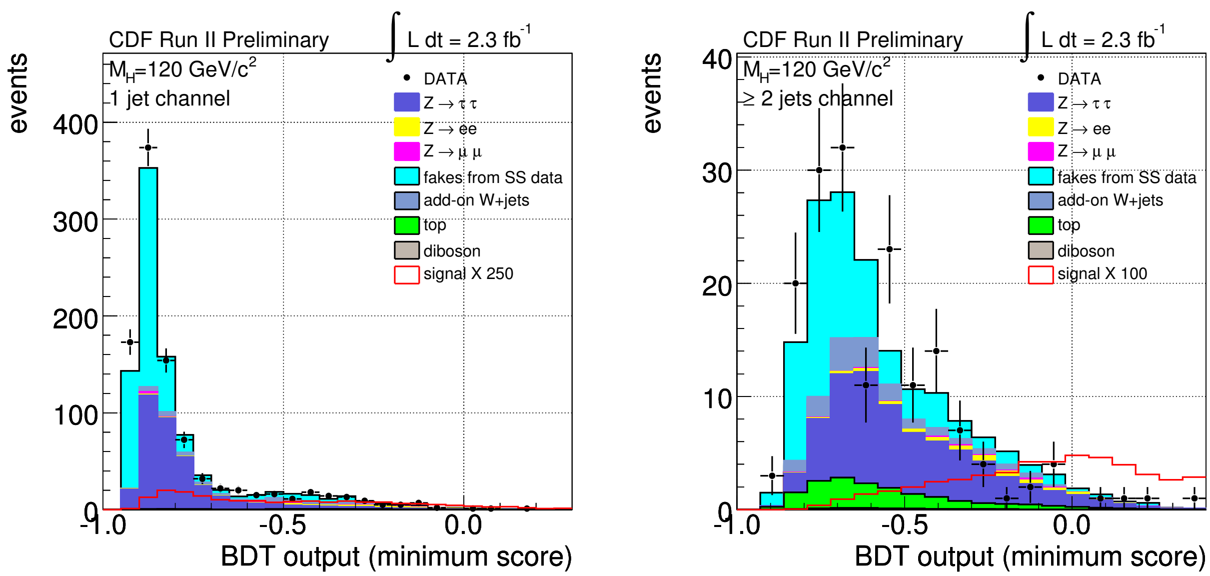
After applying the event selection, the background expectation is still significantly larger then the expected Higgs boson signal.

In order to increase the search sensitivity, we employ a multivariate technique, based on the BDT method, which allows to exploit all the event information by combining the discriminating power of different kinematic and topological variables into one single output variable, useful to extract the small signal from data.

In this search we consider the two signal channels (1 jet and $\geq$ 2 jets) separately and we train several independent BDTs, each of them optimized to get the best separation between a realistic mixture of all Higgs signal processes and one of the three dominant sources of background which appear in the event selection: $Z/\gamma* \rightarrow \tau\tau$, QCD and $t\bar{t}$, the latter only for the $\geq$ 2 jets channel.

The procedure is repeated for each Higgs mass, ranging from 110 GeV/c^2 to 150 GeV/c^2, in steps of 10 GeV/c^2. The best results in terms of sensitivity are achieved when we define the final discriminant distribution as a combination of all the BDT outputs, by keeping for each event the minimum score among the different BDTs.

The final 1 jet and $\geq$ 2 jets BDT score distributions, for the Higgs mass hypothesis of 120 GeV/c^2, are shown in figure 7 and in figure 8.



(a) 1 jet signal channel

(b) $\geq$ 2 jets signal channels

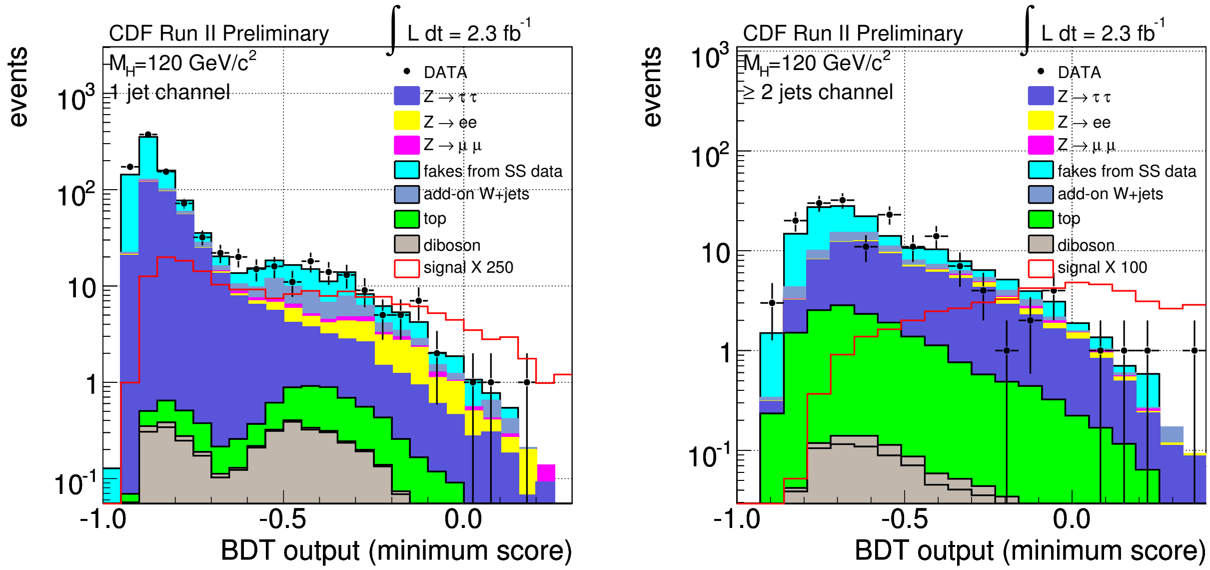
FIG. 7: Final discriminant templates for the Higgs mass hypothesis of 120 GeV/c^2.

IX. SYSTEMATIC UNCERTAINTIES

Several sources of systematic uncertainties, which can influence both the expected event yield and the shape of the distributions, affect the sensitivity of this search and have to be appropriately accounted for, by propagating their effect to the final discriminant distributions.

Since our event selection is characterized by the number of reconstruted jets in the final state, the analysis turns out to be particularly sensitive to the uncertainty assigned to the jet energy scale and has been evaluated by applying $\pm 1\ \sigma$ variations to the default jet energy correction factors. In particular, a change of the absolute corrected energy of the jets (selected by a lower threshold of 20 GeV on their transverse energy), could lead to event migrations between the different jet multiplicity channels, and then affects the kinematic acceptances of each signal and MC-derived background process.

The uncertainties on the WW/WZ/ZZ and $t\bar{t}$ cross sections are assigned to be 6% [11][12] and 10% [13] respectively. For Drell-Yan processes we take a value of 2.2%, related to previous CDF measurement uncertainties [17]. The four



(a) 1 jet signal channel

(b) $\geq$ 2 jets signal channels

FIG. 8: Final discriminant templates in logarithmic scale for the Higgs mass hypothesis of 120 GeV/c^2.

Higgs production mechanisms under consideration have different theoretical cross sections uncertainties: 10% for vector boson fusion, 5% for the associated productions and 23.5%(67.5%) for the gluon fusion in the 1 jet and $\geq$ 2 jets channels, respectively [14].

We assign an additional acceptance uncertainty of 2.3% to the Drell-Yan background, which has been estimated from the difference between the acceptance derived with ALPGEN and PYTHIA generators.

Systematics due to the modeling of initial (ISR) and final state radiation (FSR) have been evaluated for the Higgs signal processes, by considering the acceptance variations in dedicated MC samples, where the strength of ISR and FSR, in the parton showering processes, are increased or reduced by 1 σ: the major effect, up to 15%, is for the gluon fusion case.

Uncertainties related to the parton distribution function modeling are evaluated by applying a re-weight procedure on the MC samples and by comparing the default CTEQ5L model with several alternative sets of PDFs (MRST72, MRST75 and 20 orthogonal CTEQ6M). We quote the values which have been calculated in [18].

Several other common CDF analyses' systematics related to MC predictions (electron and muon identification efficiency, trigger efficiency, luminosity measurements) are all incorporated within the uncertainty on the tau ID scale factor, which is obtained by scaling MC expectations to data in the $Z/\gamma* \rightarrow \tau\tau$ control region: when this scale factor is applied, these sources cancel out. At the same time new kinds of systematics appear and need to be accurately propagated throughout the analysis: the main contributions come from the number of observed OS and SS events, the estimated W+jets rate, the Drell-Yan cross section and acceptance uncertainties.

Concerning the data-derived background, we assign a 10% of uncertainty to the fake contribution estimated from SS data and a 5%, 18% and 30% to the additional W+jets events coming from OS/SS asymmetry, for the 0 jet, 1 jet and $\geq$ 2 jets channels respectively.

The complete set of systematics uncertainties is summarized in table III.

X. RESULTS

The observed number of events in the two signal channels is consitent with the standard model background expectations: since no evidence for a Higgs boson signal is found, the discriminant distributions defined, for each channel, as a combination of the scores of the trained BDTs, are used as templates to set bayesian 95% C.L. upper limits on the SM Higgs production.

All systematic uncertainties are appropriately treated in the limit calculation, by taking into account all the possible

Systematic uncertainties (%)											
		Background					Signal				
Source		Z/→ ll	$t\bar{t}$	diboson	fakes from SS	W+jets	ggH	WH	ZH	VBF	
JES	(0 jet)	-0.6	-19.0	-0.9							
	(1 jet)	+6.2	-7.7	+7.1			+5.1	-4.8	-5.3	-3.7	
	(≥ 2 jets)	+14.2	+3.2	+11.7			+13.2	+5.4	+4.8	+5.2	
Cross section	(0 jet)	+2.2	+10.0	+6.0							
	(1 jet)	+2.2	+10.0	+6.0			+23.5	+5	+5	+10	
	(≥ 2 jets)	+2.2	+10.0	+6.0			+67.5	+5	+5	+10	
PDF		+1.0	+1.0	+1.0			+4.9	+1.2	+0.9	+2.2	
ISR	(1 jet)						+13.0	-6.1	-1.7	-2.9	
	(≥ 2 jets)						+15.5	-1.5	+0.1	-2.7	
FSR	(1 jet)						-5.0	+4.3	+1.0	+1.7	
	(≥ 2 jets)						-5.2	-2.1	+0.4	-1.1	
SS data					+10.0						
W+jets scale	(0 jet)					+5.0					
	(1 jet)					+18.0					
	(≥ 2 jets)					+30.0					
Acc.(DY)		+2.3									
tau ID SF:											
$N_{\rm obs}$		+2.8	+2.8	+2.8			+2.8	+2.8	+2.8	+2.8	
$N_{\rm SSdata}$		-3.3	-3.3	-3.3			-3.3	-3.3	-3.3	-3.3	
$N_{\rm W+jets}$		-0.3	-0.3	-0.3			-0.3	-0.3	-0.3	-0.3	
cross section(DY)		-2.1	-2.1	-2.1			-2.1	-2.1	-2.1	-2.1	
Acc.(DY)		-2.2	-2.2	-2.2			-2.2	-2.2	-2.2	-2.2	

TABLE III: Systematic uncertainties on the background and signal, expressed in %.

correlations and dependencies: a set of 10000 MC background-only pseudoexperiments, generated by varying the different process expectations accordingly to the assigned uncertainties, provide the 95% C.L. limit. The two channels are combined and the results, for the expected and the observed limits are shown in figure 9 and summarized in table IV, for the different Higgs boson masses.

Higgs Mass	Expected limit/σ(SM)					Observed limit/σ(SM)
GeV/c^2	-2 σ	-1 σ	median	+ 1 σ	+ 2 σ	
100	15.2	20.1	28.2	40.4	56.6	37.6
105	13.4	17.8	25.2	35.2	48.0	34.5
110	13.1	17.2	23.9	34.2	48.2	34.8
115	13.8	17.7	24.5	35.4	50.2	27.9
120	13.1	17.0	23.4	33.2	46.8	27.2
125	14.5	18.8	26.5	37.8	52.6	25.3
130	15.5	20.3	28.1	40.3	56.9	30.0
135	18.4	23.8	33.7	48.2	65.6	30.3
140	22.6	29.6	41.2	58.6	81.2	38.2
145	31.2	41.0	57.4	81.9	114.8	46.3
150	45.1	59.1	82.6	118.0	166.5	67.0

TABLE IV: Expected and observed 95% C.L. limit for each mass point.

XI. CONCLUSIONS

We have presented a search for the Standard Model Higgs decaying into a tau pair with an integrated luminosity of 2.3 fb^{-1} in events selected by requiring one lepton (electron or muon), one hadronically decaying tau and at least one jet.

We expect 921.7±48.9 background events in the 1 jet channel and 159.4±11.6 in the ≥ 2 jets channel, while we observe 965 and 166 events, respectively.

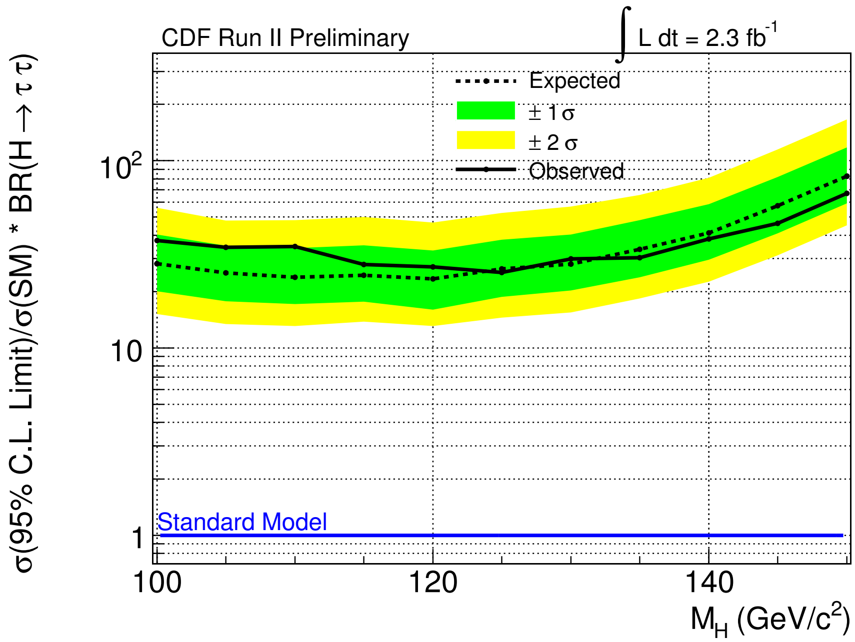


FIG. 9: Expected and observed limit as a function of the Higgs mass.

The limit at 95% C.L. on the cross section times the branching ratio of the Higgs boson decaying to tau pairs has been computed in the 100-150 GeV/c^2 mass range, in steps of 5 GeV/c^2. The value of the expected(observed) σ/σ_{SM} limit for an Higgs mass of 120 GeV/c^2 is 23.4 (27.2).

Acknowledgments

We thank the Fermilab staff and the technical staffs of the participating institutions for their vital contributions. This work was supported by the U.S. Department of Energy and National Science Foundation; the Italian Istituto Nazionale di Fisica Nucleare; the Ministry of Education, Culture, Sports, Science and Technology of Japan; the Natural Sciences and Engineering Research Council of Canada; the National Science Council of the Republic of China; the Swiss National Science Foundation; the A.P. Sloan Foundation; the Bundesministerium fuer Bildung und Forschung, Germany; the Korean Science and Engineering Foundation and the Korean Research Foundation; the Particle Physics and Astronomy Research Council and the Royal Society, UK; the Russian Foundation for Basic Research; the Comision Interministerial de Ciencia y Tecnologia, Spain; and in part by the European Community's Human Potential Programme under contract HPRN-CT-20002, Probe for New Physics.

---

[1] The LEP Electroweak Working Group, `http://lepewwg.web.cern.ch/LEPEWWG/`. 2
[2] A. Hoecker *et al.*, Proc. Sci **ACAT2007**, 040 (2007); L. Breiman *et al.*, *Classification and Regression Trees*, Wadsworth and Brooks, Monterey, CA, 1984. 2
[3] R. Blair *et al.*, The CDF Collaboration, FERMILAB-PUB-96/390-E (1996). 2
[4] A. Sill *et al.*, Nucl. Instrum. Methods A **447**, 1 (2000). 2
[5] A. Affolder *et al.*, Nucl. Instrum. Methods A **453**, 84 (2000). 2
[6] T. Affolder *et al.*, Nucl. Instrum. Methods A **526**, 249 (2004). 2
[7] R. Brun *et al.*, CERN-DD-78-2-REV, CERN-DD-78-2 4
[8] T. Sjostrand, S. Mrenna, P. Skands, Peter, JHEP, **05** 026 (2006) 4
[9] M.L. Mangano *et al.*, J. High Energy Phys. **0307** 001 (2003) 4
[10] H.L. Lai *et al.*, CTEQ Collaboration, Eur. Phys. J. **C12** 375-392 (2000) 4
[11] J.M. Campbell and R.K. Ellis, http://mcfm.fnal.gov 11
[12] J.M. Campbell and R.K. Ellis, Phys. Rev. **D60**, 113006 (1999) 11
[13] U.Langenfeld,S.Moch and P.Uwer, Phys. Rev. **D80**, 054009 (2009) 11

[14] Tev4LHC Working Group, `http://maltoni.home.cern.ch/maltoni/TeV4LHC/SM.html` 12
[16] A. Bhatti *et al.*, Nucl. Instrum. Meth. **A566**, 375 (2006) 4
[16] A. Bhatti *et al.*, Nucl. Instrum. Meth. **A566**, 375 (2006) 4
[17] D. Acosta *et al.*, Phys. Rev.Lett. **94**, 091803 (2005) 11
[18] The CDF Collaboration, Search for the Standard Model Higgs Boson in H→ $\tau\tau$ Channel at CDF Run II. Simultaneous Search for WH/ZH/VBF/H Processes, CDF Public Note 9248, 2008 12